This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2019.2953717, IEEE Transactions on Medical Imaging

De-smokeGCN: Generative Cooperative Networks for Joint Surgical Smoke Detection and Removal

Long Chen, Wen Tang, Nigel W. John, Tao Ruan Wan, and Jian Jun Zhang

***Abstract*—Surgical smoke removal algorithms can improve the quality of intra-operative imaging and reduce hazards in image-guided surgery, a highly desirable post-process for many clinical applications. These algorithms also enable effective computer vision tasks for future robotic surgery. In this paper, we present a new unsupervised learning framework for high-quality pixel-wise smoke detection and removal. One of the well recognized grand challenges in using convolutional neural networks (CNNs) for medical image processing is to obtain intra-operative medical imaging datasets for network training and validation, but availability and quality of these datasets are scarce. Our novel training framework does not require ground-truth image pairs. Instead, it learns purely from computer-generated simulation images. This approach opens up new avenues and bridges a substantial gap between conventional non-learning based methods and which requiring prior knowledge gained from extensive training datasets. Inspired by the Generative Adversarial Network (GAN), we have developed a novel generative-collaborative learning scheme that decomposes the de-smoke process into two separate tasks: smoke detection and smoke removal. The detection network is used as prior knowledge, and also as a loss function to maximize its support for training of the smoke removal network. Quantitative and qualitative studies show that the proposed training framework outperforms the state-of-the-art de-smoking approaches including the latest GAN framework (such as PIX2PIX). Although trained on synthetic images, experimental results on clinical images have proved the effectiveness of the proposed network for detecting and removing surgical smoke on both simulated and real-world laparoscopic images.**

***Index Terms*—Endoscopy, Image enhancement, Machine learning, De-smoking.**

I. Introduction

Surgical smoke is a by-product produced by energy-generating devices during surgery. Surgical smoke in intra-operative imaging and image-guided surgery [1] can severely deteriorate the image quality [2] and pose hazards to surgeons [3]. Thus, improving the quality of intra-operative images is highly desirable in many clinical applications. Surgical smoke also poses significant issues [4] in future advanced medical imaging tasks such as robotic surgery, real-time surgical reconstruction and augmented reality, in which the effectiveness of computer vision is pertinent.

Although smoke evacuation devices are available for smoke removal, these devices are unsuitable for image-guided surgery. Methods published recently are mainly based on conventional image processing algorithms, which have taken a two-steps approach: filtering out smoke first, then recovering images as sharply and clearly as possible [5] [6] [7] [8] [9]. These two-steps based approaches suffer from the problem of fidelity loss due to image over-enhancement. More recently introduced end-to-end deep learning approaches [10] for surgical de-hazing and de-smoking start to emerge. Although there have been some promising results, challenging issues must be solved before the methods can be introduced into medical practice:

- Large amounts of intra-operative datasets are difficult to collect and scarcely available for CNNs to learn implicit de-smoking functions, especially for learning surgical scenes.
- There is a danger of overfitting learning-based methods to limited amount and variations of training data, leading to poor performance when tested on real-world data.
- Sometimes smoke is also an important signal of the ablation process. Removing the smoke can have a reverse effect if the process is not quantifiable and controllable.

In this paper, we formulate tasks of smoke detection and removal as two joint learning processes and propose a novel computational framework for *unsupervised collaborative learning*. Our well-designed CNNs learn the smoke detection and removal from rendering smoke on laparoscopic videos. In summary, contributions of this work include:

- Novel integration of a graphics rendering engine into our learning framework for continuously outpouring unlimited training data without the need for any manual labeling.
- Decomposition of the smoke removal task into two loosely-coupled sub-module tasks: pixel-level smoke detection and smoke removal based on detection results. The two loosely-coupled tasks not only prevent overfittings to the synthetic datasets, but also make the surgeon aware of how much smoke is removed.
- A novel training framework for Generative Collaborative Networks (GCN) which maximally exploits the potential of the proposed networks for smoke detection and removal.
- Compared with conventional image processing ap-

Long Chen was with the Department of Creative Technology, Bournemouth University, Poole, UK, BH12 5BB

Wen Tang, Corresponding author, is with the Department of Creative Technology, Bournemouth University, Poole, UK, BH12 5BB

Nigel W. John is with the Department of Computer Science, University of Chester, Chester, UK, CH1 4BJ

Tao Ruan Wan is with the Faculty of Science and Engineering, University of Bradford, UK, BD7 1DP

Jian Jun Zhang is with the National Centre for Computer Animation, Bournemouth University, UK, BH12 5BB

Manuscript received September 19, 2018;

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2019.2953717, IEEE Transactions on Medical Imaging

proaches, the proposed framework is capable of removing surgical smoke with a global contextual understanding and recover more realistic tissue colours.
- Compared with the latest Generative Adversarial Network (GAN), our method produces more faithful results without adding "fake" scars and surface reflections.

Through quantitative and qualitative evaluations, the results have proved that the proposed method outperforms the GAN framework and the state-of-the-art smoke removal approaches. We show that using computer-generated synthetic images the network is able to remove real surgical smoke on laparoscopic images effectively.

II. Related Work

Image processing and computer vision research communities have been tackling general image de-hazing and de-smoking tasks for decades, ranging from obtaining clear outdoor scenes affected by weather conditions to recovering surgical scenes. Typically, methods for smoke removal are either based on image processing or machine learning.

A. Atmospheric Scattering Model

One of the most classic models to describe hazy or smoky images is the atmospheric scattering model[11] [12] [13].

$$I(x) = J(x)t(x) + A(1 - t(x)) \tag{1}$$

where $I(x)$ is the observed hazy image, $J(x)$ is the clear scene to be recovered from, A is the global atmospheric light, $t(x)$ is the medium transmission, which can be described by $t(x) = e^{-\beta d(x)}$, where β is the atmosphere scattering coefficient and $d(x)$ is the distance. The atmospheric scattering model is based on the strong assumption that haze is homogeneous and the light source is at a far distance so that rays and beams (such as sunlight) are parallel. In contrast, in minimally invasive surgical scenes smoke concentration can vary greatly and light sources are close to the scene. It is very hard to predict $t(x)$. The lighting illumination is usually uneven in the scenes due to very close distances between the light source and tissues. Therefore, the general atmospheric scattering model is inappropriate for surgical applications.

B. Dark Channel Prior based De-smoking

The dark-channel prior proposed by He *et al.* [14] is a simple but effective method for predicting a transmission map based on observations of the natural property of haze-free images – pixels should have at least one colour channel with very low-intensity values. This method can cause chromatic changes and fidelity loss in minimally invasive surgical scenes, because the close-distance of the direct light source to the tissue surface produces highly-illuminated pixels such as tissue reflections and light colour on fat tissues can be falsely detected as hazy, violating the dark-channel prior assumption.

Tchaka *et al.* [8] used an adaptive dark-channel prior with a histogram equalization to remove smoke from endoscopic images. This method applied empirically chosen parameters. Although histogram equalization can enhance the colour and contrast, due to the limitation of the dark-channel prior, the original and real colours are not preserved.

C. Optimization-based De-smoking

Fattal *et al.* [15] further refined the dark-channel prior model by taking into account the surface shading in addition to the scene transmission and using a Gaussian Markov Random Field (MRF) model to recover the haze-free image. Nishino *et al.* [16] modeled the chromaticity and the depth, also with the use of a factorial MRF to obtain more accurate scene radiance estimations. Based on the observation that hazy-free images tend to have much higher contrast, Tan *et al.* [17] proposed a local contrast maximizing method, which also optimized MRF models. Meng *et al.* [18] introduced an inherent boundary constraint on the transmission function to recover more image details and structures. Baid *et al.* [6] presented an unified Bayesian formulation for simultaneously de-smoking, specularity removal and de-noising in laparoscopy images. This method proposed several priors via probabilistic graphical models and sparse dictionaries to model colours and textures of un-corrupted images. A variational Bayes Expectation Maximization optimization was used to minimize the overall energy function and infer un-corrupted images from corrupted images.

Global-contextual awareness is the key feature of the proposed method in this paper. Despite well-designed MRFs priors, these hand-crafted prior models have a limited expressive power and lack global contextual understandings of ill-posed problems like surgical de-smoking. Another common weakness is that these methods were all trying to minimize prior features that tend to be hazy, which usually lead to over-enhanced image colours and contrasts and also suffer from fidelity loss.

D. Learning based De-smoking

With the recent success of deep learning algorithms, many deep learning frameworks are introduced to solve de-hazing and de-smoking problems. DehazeNet [10] is an end-to-end learning system for haze removal in single images by learning a medium transmission map that is subsequently used to recover a haze-free image through the atmospheric scattering model. AOD-Net [19] also integrates the atmospheric scattering model into its network structure and achieves an all-in-one and end-to-end training. As described above, networks based on the Atmospheric Scattering Model are not suitable for surgical scenes. Furthermore, these network structures are also very shallow for learning and recovering fine image details.

E. Novelty Compared to Previous Work

Most of the above works rely on Equation 1 (the atmospheric scattering model) to solve the de-hazing problem. However, in minimally invasive surgical scenes, smoke is often non-uniform and light beams are usually nonparallel and uneven, making the problem ill-posed. In our previous paper [20], we proposed to use an U-Net structure to remove surgical smoke. Although it works well on synthetic datasets, the end-to-end training will be like to overfit to the original datasets and perform poorly on real datasets. Wang *et al.* [21] proposed a multi-scale learning based de-smoking method that

uses Laplacian image pyramids as extra information to train a de-smoking network. In this paper, we reformulate Equation 1 as fully end-to-end learning processes by firstly estimating the smoke mask, then use it as the prior knowledge for another neural network to learn the ill-posed smoke removal function. The proposed method not only achieves better results, but also reduces the over-fitting and makes the network more robust to deal with real-world images. The pixel-level smoke detection results can also lead to many useful applications such as estimating smoke volumes and improving contextual understandings of surgical smoke.

III. Methods

The goal of removing smoke is a straightforward one – we want to remove the smoke while maximally keeping the original colours of non-smoke areas. We decompose the smoke removal task into two sub-tasks: smoke detection and smoke removal. Two fully connected convolutional networks are used to learn the smoke detection and removal tasks separately but also cooperatively:

- The smoke detection network focuses on detecting smoke and providing a pixel-level smoke mask.
- The smoke removal network focuses on removing smoke based on the smoke mask and smoke images.
- The smoke detection network serves as supervision to examing the smoke removal result and provides gradients for optimizing the smoke removal network.

As shown in Figure 1, the proposed training pipeline consists four main parts: Smoke Synthesis (1); Smoke Detection (2); Smoke Removal (3); and Detection-after-generation (DaG) supervision (4). Each of these components is detailed below.

A. Smoke Synthesis

Making large datasets available for training neural networks is an extremely costly and time-consuming undertaking, especially as medical datasets not only take up valuable medical resources, but also require great accuracy and quantities to satisfy the medical practice standard. Tasks of smoke detection and removal are more difficult since image pairs (with and without the presence of smoke) and the smoke density mask are required. It is nearly impossible to acquire these image pairs and density masks through manual labeling.

To tackle this problem, we employ a modern 3D graphics rendering engine for continuously rendering smoke onto laparoscopic images to generate smoked images. In doing so, we can also obtain smoke masks to train the pixel level smoke detection and removal tasks. We use an open source 3D creation software [1] for the synthesis of smoke images for training. Advantages of using a standard rendering engine, instead of employing a physically-based haze formation model as in [10] [22] or a Perlin noise function [23] to generate smoke procedurally, are two-fold. Firstly, in laparoscopic scenes, surgical smoke is often generated locally and is independent to the depth, so there is no reason to use a traditional haze model for rendering surgical smoke. Secondly, nowadays modern graphics rendering engines can produce more realistic smoke shapes and density variations based on well-developed built-in models, which are also physically-based.

Real laparoscopic images available from the Hamlyn Centre Laparoscopic / Endoscopic Video datasets [2] [24] and Cholec80 dataset [3] [25] are used as background images. The variance of the Laplacian [26] is firstly used for screening images, and a second-round manual inspection ensures the images contain no presence of surgical smoke for ground truth. A total of 21,000 images are sampled from 91 videos as the smoke-free source images.

The smoke I_{smoke} is rendered by our render engine with local colours and transparencies and positions controlled by input parameters of random intensity T_{rand}, density D_{rand} and position P_{rand}.

$$I_{smoke}(x,y) = Blender(T_{rand}, D_{rand}, P_{rand}) \tag{2}$$

The randomly generated smoke I_{smoke} is then overlaid onto each of the background images $I_{smoke-free}$ to composite smoked surgical images $I_{smoked-image}$.

$$I_{smoked-image}(x,y) = I_{smoke-free}(x,y) + I_{smoke} \tag{3}$$

The smoke mask I_{mask} is derived from the luminosity of R, G, B channels from the rendered smoke I_{mask}

$$\begin{aligned} I_{mask}(x,y) = (0.3 * I_{smoke}(x,y)^R) + \\ (0.59 * I_{smoke}(x,y)^G) + (0.11 * I_{smoke}(x,y)^B) \end{aligned} \tag{4}$$

The variations of the rendered smoke ensure that there is no over-fitting for the network to certain smoke intensities, densities and locations. With the help of a powerful rendering engine, we are able to synthesize an unlimited amount of realistic images with simulated surgical smoke for network training.

B. Smoke Detection

We use a smoke detection network to generate the pixel-wise smoke density. Benefits of such an approach are:

- The smoke detection provides a pixel-level smoke density to provide information about the amount and the position of the surgical smoke.
- The detected smoke serves as the prior information fed into the subsequent smoke removal network.
- The smoke removal network is optimized under the supervision of the smoke detection network. (see Section III-D)

We employ an U-Net [27] based fully convolutional encoder-decoder network structure with parameters θ_d for pixel level smoke detection: $D(I_{smoked-image}) \underset{\theta_d}{\rightarrow} I_{mask}$

As shown in Fig. 3, the smoke detection network consists of four convolutional layers as an encoder to abstract the input image efficiently into a high-dimensional feature tensor that is $1/2^4$ the original size and with 512 channels. For the decoder, four symmetrical de-convolutional layers are used to recover the feature tensor into a full original sized smoke mask. Each

[1] https://www.blender.org/

[2] http://hamlyn.doc.ic.ac.uk/vision/

[3] http://camma.u-strasbg.fr/datasets

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2019.2953717, IEEE Transactions on Medical Imaging

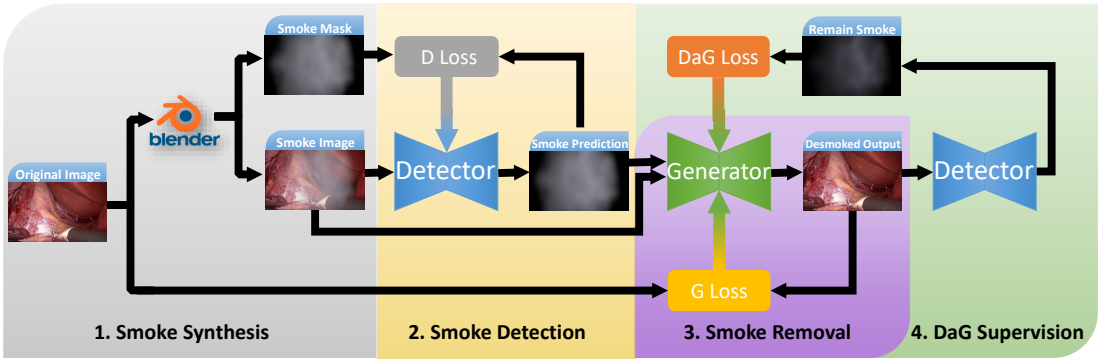

Fig. 1. Overview of our framework for unsupervised learning of smoke removal

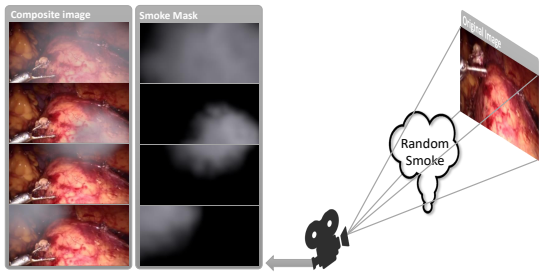

Fig. 2. Left: Rendered images and smoke masks. Right: A 3D illustration of the rendering process.

layer is with a kernel size four and a stride size two, followed by leaky ReLU layers and a batch normalization. Skip layers are connected with the corresponding layer pairs from encoder and decoder for preserving the high-level information to ensure high-quality per-pixel smoke detection after up-sampling.

Reasons for using a shallow network with fewer layers are:

- The intended smoke detection is a simple task compared with that of smoke removal, so a shallow network is sufficient.
- A shallow network will have fewer weights to prevent network over-fitting to specific smoke patterns.
- A shallow network will accelerate the speed of training and inferring.

The loss function for the smoke detection network is:

$$\begin{aligned}\mathcal{L}_D^{total} = \sum_{x,y}(\alpha_d \underbrace{\left|\widehat{I}_{mask}(x,y) - I_{mask}(x,y)\right|}_{L1\ loss} \\ + \underbrace{\beta_d \left|\widehat{I}_{mask}(x+1,y) - \widehat{I}_{mask}(x,y)\right|}_{x\ smooth\ term} \\ + \underbrace{\beta_d \left|\widehat{I}_{mask}(x,y+1) - \widehat{I}_{mask}(x,y)\right|}_{y\ smooth\ term})\end{aligned} \tag{5}$$

where $\widehat{I}_{mask}(x,y)$ and $I(x,y)_{mask}$ are estimated smoke mask and ground truth smoke mask. We use a combination of a L1 loss and two smoothness terms for the total loss of the network. We take the L1 norms of the predict smoke masks' gradients along x and y directions as smoothness terms. Due to the fact that smoke tends to be smooth, applying penalties on smoke masks' discontinuities can ensure the accurate, smoothness and realism of smoke mask predictions.

C. Smoke Removal

The smoke mask I_{mask} estimated by the smoke detection network is further used as prior knowledge for learning smoke removal. As can be seen from the second network in Figure 3, the smoke mask I_{mask} and the smoke image $I_{smoked-image}$ are concatenated into a 4-channel layer before the input into the smoke removal network G with parameters θ_g.

$$G(I_{mask} \oplus I_{smoked-image}) \underset{\theta_g}{\rightarrow} I_{smoke-free} \tag{6}$$

An encoder-decoder network similar to the smoke detection network is used for generating smoke-free images. A deeper network with eight convolutional layers for the encoder is used to compress the input image into a 512 channel feature tensor, and eight de-convolutional layers to recover it into a full-size smoke-free mask. To prevent the loss of image details, following the U-Net structure [27], skip connections are implemented to transfer high-level information directly to the bottom of the network. We use a doubled number of layers for learning smoke removal since it is an ill-posed problem that requires contextual understandings of the image to recover the correct colours of the smoked regions.

The first part of the loss function of the smoke removal network is a L1 loss between the estimated smoke-free image and the original smoke-free image without the simulated smoke:

$$\mathcal{L}_G^{L1} = \sum_{x,y}\left|\widehat{I}_{smoke-free}(x,y) - I_{smoke-free}(x,y)\right| \tag{7}$$

D. Detection after Generator (DaG) Supervision

To take full advantage of the proposed smoke detection network, we guide the smoke removal process further by using the smoke detection network as the second supervision stage. The estimated smoke-free image $\widehat{I}_{smoke-free}$ is fed into the smoke detection network after generated from the smoke removal network:

$$D(\widehat{I}_{smoke-free}) \underset{\theta_d}{\rightarrow} 0 \tag{8}$$

To make sure the smoke removal network G works cleanly (there is no smoke left after the removal), the goal is to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2019.2953717, IEEE Transactions on Medical Imaging

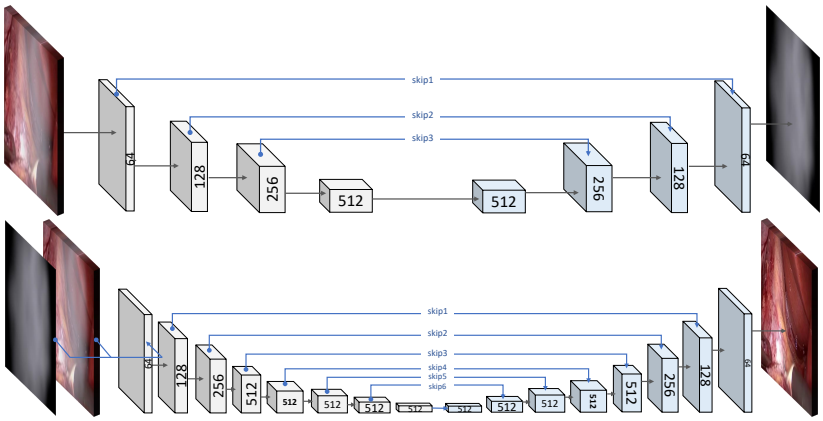

Fig. 3. Network structures of the smoke detection network (top) and the smoke removal network (bottom)

minimize the output of the detected smoke to provide gradients for the smoke removal network G. Therefore, the second part of the loss function is L1 norm of the predicted smoke mask based on the predicted smoke-free image, which can also be expressed as L1 norm of the detector after the generator:

$$\begin{aligned}\mathcal{L}_G^{DaG} &= \sum_{x,y} \left| D(\widehat{I}_{smoke-free}(x,y)) \right| \\ &= \sum_{x,y} |D(G(I_{smoked-image}(x,y)))|\end{aligned} \tag{9}$$

The total loss of the smoke removal network is:

$$\mathcal{L}_G^{total} = \alpha_g \mathcal{L}_G^{L1} + \beta_g \mathcal{L}_G^{DaG} \tag{10}$$

where α_g and β_g are weights for $L1$ loss and DaG loss.

IV. Experiments

This section describes the experimental setup and evaluation results of the proposed smoke detection and removal networks. We provide quantitative and qualitative comparisons of our results with eleven state-of-the-art approaches.

A. Implementation details

The proposed networks are implemented in Tensorflow and trained on a workstation with an NVIDIA Titan X GPU (12G Graphic Memory).

For training, we apply gradient descent steps of D and G separately to avoid interference between each other. The D is firstly trained for 1 epoch, so that the D can roughly provide a smoke mask. After this process, the D and G are trained iteratively. When training G, the network parameters D are frozen. An Adam solver is used for training with the following hyper parameters: learning rate 0.0002, and momentum parameters $\beta_1 = 0.5$, $\beta_2 = 0.999$; batch size of 16. We empirically set the weights $\alpha_d = \beta_d = 1$, $\alpha_g = 1$, $\beta_g = 100$ based on several tests. In our implementation, a drop-out is used in the 5$^{\text{th}}$ layer for the smoke detection network and the 9$^{\text{th}}$ layer for the smoke removal network with a change of 50% to prevent over-fitting.

For the training dataset, we sampled 21,000 images without the presence of surgical smoke amongst 91 videos from the Hamlyn Centre Laparoscopic / Endoscopic Video datasets [4] [24] and Cholec80 dataset [5] [25]. The method described in Section III-A was used to produce $I_{smoked-image}$ and I_{mask}

To fulfil the leave-patients/videos-out criteria, for the testing dataset, we sampled 1,228 smoke-free images from 27 cholecystectomy procedure videos in the m2cai16-workflow dataset [28] [29]. The same procedures were applied to the testing images to produce $I_{smoked-image}$ for testing dataset.

All images are re-sized to 256x256 pixels for efficient training and testing. The training time is around 14 hours. When in testing mode, the networks can estimate smoke masks and smoke-free images at 45 fps.

B. Comparison methods

For quantitative evaluations, we report evaluation criteria in terms of the difference between the pair of smoke-free images and de-smoked results, including the Mean Squared Error (**MSE**), the Peak Signal-to-Noise Ratio (**PSNR** in dB) and the Structural Similarity Index (**SSIM**). The lower **MSE**, the higher **PSNR**, and **SSIM** indicate that the estimated smoke-free images are similar to the real smoke-free images, which means a better de-smoking capability.

The proposed method is compared with eleven state-of-the-art de-smoking and de-haze methods including both conventional image processing methods and the latest deep learning based methods. Theses include Dark Channel Prior (**DCP**) [14], Boundary Constraint and Contextual Regularization (**BCCR**) [18], Fusion-based Variational Image Dehazing (**FVID**) [30], Automatic Recovery of Atmospheric Light (**ATM**) [31], colour Attenuation Prior (**CAP**) [32], DEnsity of Fog Assessment based DEfogger (**DEFADE**) [33], Enhanced Variational Image Dehazing (**EVID**) [34], Non-local Image Dehazing (**NLD**) [35], Graphical Models and Bayesian Inference (**GMBI**) [6], and deep learning based methods including the All-in-One Dehazing Network (**AOD-NET**) [19], Image-to-Image Translation with Conditional Adversarial Networks (**PIX2PIX**) [36]. All of the source codes were collected from the author or third-party implementations, using the default

[4] http://hamlyn.doc.ic.ac.uk/vision/
[5] http://camma.u-strasbg.fr/datasets

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2019.2953717, IEEE Transactions on Medical Imaging

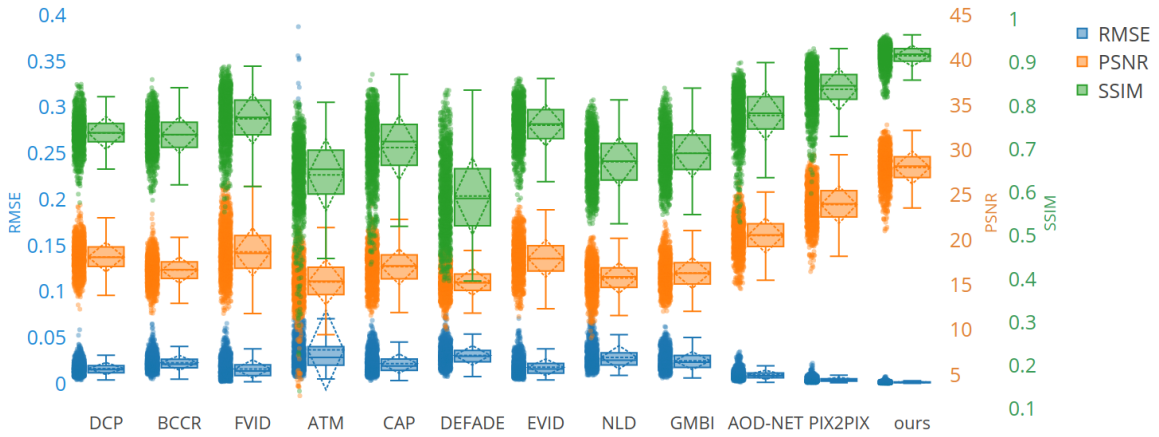

Fig. 4. Box plots of the 3 metrics MSE, PSNR and SSIM for our results and 11 previous approaches.

TABLE I
QUANTITATIVE RESULTS

Method	Ref	DL?	Platform	Lower is better MSE	Higher is better PSNR	Higher is better SSIM	Time/frame
DCP	[14]	No	Matlab	0.016 ± 0.006	18.117 ± 1.641	0.738 ± 0.038	3.612
BCCR	[18]	No	Matlab	0.023 ± 0.008	16.664 ± 1.462	0.733 ± 0.042	0.275
FVID	[30]	No	C/Matlab	0.016 ± 0.010	18.694 ± 2.669	0.770 ± 0.058	5.360
ATM	[31]	No	Matlab	0.037 ± 0.043	15.327 ± 2.518	0.641 ± 0.084	21.508
CAP	[32]	No	Matlab	0.022 ± 0.010	17.036 ± 1.976	0.704 ± 0.074	0.118
DEFADE	[33]	No	Matlab	0.031 ± 0.011	15.353 ± 1.471	0.592 ± 0.089	2.123
EVID	[34]	No	C/Matlab	0.018 ± 0.009	17.955 ± 2.074	0.756 ± 0.048	5.806
NLD	[35]	No	Matlab	0.029 ± 0.013	15.779 ± 1.689	0.671 ± 0.056	5.016
GMBI	[6]	No	Matlab	0.025 ± 0.010	16.338 ± 1.699	0.691 ± 0.056	2.210
AOD-NET	[19]	Yes	Caffe	0.010 ± 0.005	20.509 ± 1.931	0.778 ± 0.057	0.017
PIX2PIX	[36]	Yes	Tensorflow	0.005 ± 0.002	23.938 ± 2.069	0.839 ± 0.049	**0.010**
Ours(G Only)	-	Yes	Tensorflow	0.003 ± 0.001	26.590± 1.876	0.902 ± 0.025	0.012
Ours	-	Yes	Tensorflow	**0.002 ± 0.001**	**28.059± 1.820**	**0.916 ± 0.024**	0.022

parameters specified in their papers. It is worth noting that for DL-based methods [19] [36], we trained our networks with the same datasets and the same number of epochs for a fair comparative study.

C. Evaluation on Testing dataset

The testing dataset for the comparative study and the evaluation of our trained model contains 1,228 images. As can be seen from the box plots in Fig. 4 and Table IV-A, our method outperforms all of the previous de-hazing and de-smoking methods in terms of MSE, PSNR and SSIM, with very small standard deviations, indicating the robustness of our proposed system. We also report the average computational time for all of the compared state-of-the-art methods in the last row. It can be seen that deep learning (DL) based methods take significantly less time to estimate smoke-free images compared with conventional image processing methods. It is worth noting that, as our framework is a series-connection of two networks when testing, the computation time is doubled compared to the single network approaches, but can still run in 1.5x real-time at 45 fps.

As shown in Fig. 5, we display six sets of smoke-free images $I_{smoke-free}$, smoke masks I_{mask}, rendered smoke images $I_{smoked-image}$ (the only input to all methods), de-smoked results of the eleven previous methods and the output of our method $\widehat{I}_{smoke-free}$, as well as the estimated smoke mask $\widehat{I}_{mask}$. We found that most of the previous approaches can only effectively remove smoke to a certain degree, of which DCP seems to be the best one amongst non-deep learning methods. But there are still many problems for the non-deep learning methods, including:

- Not robust enough to smoke variations (position, density, and texture) and can produce unstable results (eg. ATM).
- Cannot recover correct colours for smoke-covered areas.
- Colour shift for non-smoke areas.
- Suffer from over-saturated (eg. DCP, BCCR, DEFADE) or under-saturated (eg. ATM EVID, GMBI) problems.

In contrast, our method can totally overcome these problems. The proposed method can not only focus on smoke-covered areas and retain smoke-free areas but also recover the correct tissue colours based on the contextual knowledge learned by the network. However, it is still worth noting that the non-learning based smoke removal methods often involve parametric models, which are usually not tuned for medical images but natural images. It is still interesting to see how well these methods perform in medical images.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2019.2953717, IEEE Transactions on Medical Imaging

Fig. 5. Qualitative results on synthetic testing dataset. The 1st, 2nd and 3rd row of the image matrix demonstrate the smoke-free images, rendered smoke masks and simulated smoke images. 4th-15th row show the de-smoking results from previous and our methods. Our estimated smoke mask is shown in the last column.

To prove that our smoke mask as a prior will improve the smoke removal result, we added an ablation study that is marked as "Ours(G Only)", which is the "generator only" version of our network without the smoke mask as a prior. Although smoke mask as a prior only marginally improves the quantitative result on simulated test data, the more important meaning of the smoke mask as a prior lies in its generalization ability that can overcome the overfitting to synthetic smoke. The substantial improvement achieved on real data also proves this point.

The result of AOD-NET is only slightly above the conventional image processing based methods and worse than the U-Net (our Generator only version), although it is a learning-based method trained on our training dataset. This could be due to multiple reasons: 1) The AOD-NET is still based on the Atmospheric Scattering Model, and as discussed in Section II A the Atmospheric Scattering Model does not lend itself to surgical applications due to the complex lighting conditions and smoke being heterogeneous. 2) The AOD-NET uses a shallow CNN architecture that only has five convolutional layers, while the U-Net structure that we used has 16 layers and are separated to encoder and decoder for better abstraction. It is worth noting that GAN-based methods like PIX2PIX, due to the characteristic of the GAN loss, the network learns to add "fake" features to make the image look like a smoke-free image. However, these features are selected by the machine and totally uncontrollable. As can be seen from Fig. 5, the PIX2PIX network has learned to add fake scars and reflections to the results, which is very harmful and can influence a surgeons judgment if used during surgical interventions.

D. Smoke removal limit test

Not only structural information can be blocked by smoke but also colour information can be fade. This loss of information is usually irreversible, depending on how thick the smoke is. To further evaluate the capability of networks to recover smoke-free images under different smoke densities, we conduct a performance study of de-smoking under ten different smoke densities. We randomly selected 100 images from the 2005 test datasets and rendered 10 fixed-position smoke onto each image with different smoke densities range from 0 to 10, where 0 means no rendered smoke, 9 means the maximum smoke density.

As shown in Fig. 7, we present the rendered smoke images $I_{smoked-image}$ in the first row with 10 smoke levels, and the de-smoked results from eleven previous methods, and our method shown in the last row. The results have shown that most of the previous methods cannot recover the correct colours of the dark-red tissues in the center of the images. Also, a common problem of previous methods is that estimated smoke-free images become blurry with the increase of the smoke density. In contrast, deep learning based methods give better results because the network learns to recover the correct colours based on the contextual information. It is interesting to see that PIX2PIX has produced similar results as ours, but became un-controllable after smoke level 7 and started to add "fake" reflections on the results. Our method has produced very clean results with only a minor saturation change, which is very hard to recover under very thick smoke.

Quantitative results are shown in Fig. 6. We show curves of MSE, SSIM and PSNR between image pairs of de-smoked image and smoke-free image for our results and the eleven state -f the art methods under 10 different smoke levels. Our results yield the lowest MSE as well as the highest SSIM and PSNR for all 10 smoke levels, which significantly outperform all of the previous methods.

We also plotted curves without any de-smoking process as a baseline. We found that for most of the previous approaches the results are worse than the baseline even from the beginning with no smoke, but with the rise of the smoke levels, results become better than the baseline. This is because these approaches often result in the shift of colours, the increase of contrast and saturation, which have an impact on the error measurement over the first few smoke levels. In contrast, our method has produced very robust results to the rise of the smoke level due to our novel learning frameworks that can recover the correct tissue colours under circumstances of zero smoke as well as that of very high smoke densities.

E. Evaluation on in-vivo data

Although our networks are trained purely on synthetic smoke images, we also evaluate our network on *in-vivo*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2019.2953717, IEEE Transactions on Medical Imaging

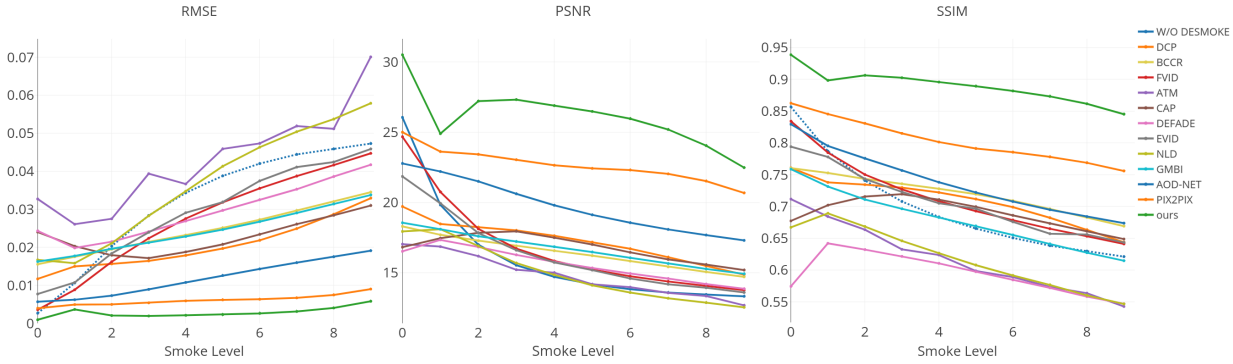

Fig. 6. The quantitative results of our smoke removal limit test. From left to right: the MSE, PSNR and SSIM results for our method and 11 comparison approaches under 10 different smoke levels

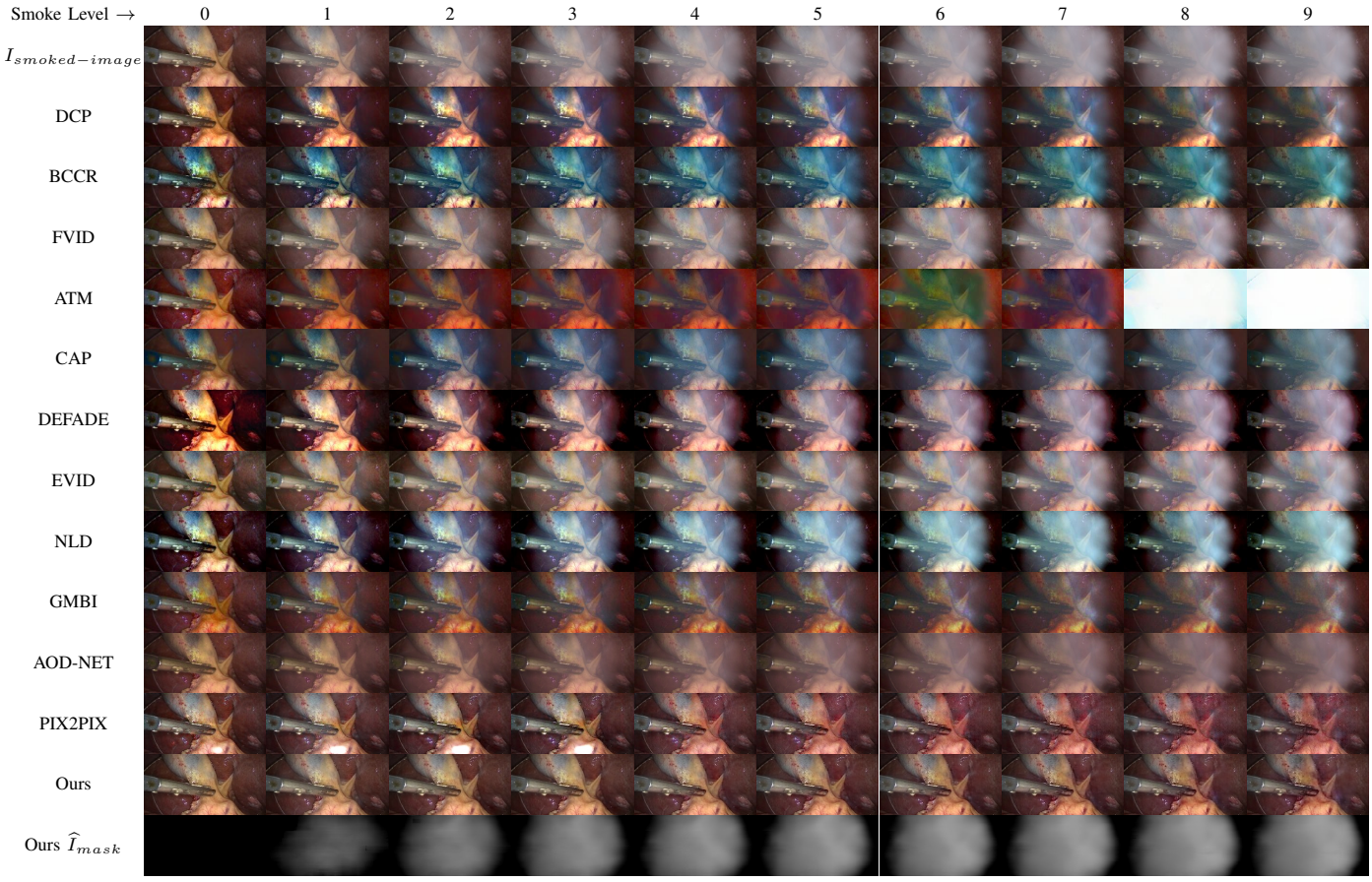

Fig. 7. Quantitative result for our smoke removal density test. Each column of the image matrix shows the de-smoking results from different methods, and each row shows the de-smoking results of different smoke level for the same method.

datasets to test the ability of our method for removing real surgical smoke. 81 images with the presence of smoke are manually picked from the Hamlyn Centre Laparoscopic / Endoscopic Video datasets and Cholec80 dataset [25] for evaluation.

Fig. 8 shows de-smoking visual results on *in-vivo* data. Again, we found that some of the previous approaches either suffer from an image over-enhancement problem (such as DCP, BCCR, ATM, DEFADE) or cannot recover clear views (such as FVID, EVID). For deep learning based methods, it appears that colours are well recovered without over-enhancement. A detailed inspection indicates that AOD-NET cannot recover clear views due to the use of a very shallow network. For PIX2PIX, there is also some smoke remaining in the result. Note that the fourth example $I_{irregular}$ is a failure case, where the smoke appeared as an irregular shape. We find that all learning based methods fail in this case, due to the fact that our simulated training data did not take the irregular shape into account. However, we believe that by applying a more aggressive random shape strategy with the simulation of training data, then this problem can be easily overcome.

To fully understand the effectiveness of our GCN training

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2019.2953717, IEEE Transactions on Medical Imaging

TABLE II
FADE SCORE ON THE *in-vivo* DATASET FROM OUR METHOD AND 11 COMPARISON APPROACHES

Method	FADE Score	
	Avg.	Std.
DCP [14]	0.4315	0.1150
BCCR [18]	0.3805	0.1147
FVAR [30]	0.8722	0.2583
ATM [31]	0.6582	1.7753
CAP [32]	0.6082	0.2481
DEFADE [33]	0.6285	0.3993
EVAR [34]	0.5383	0.1409
NLD [35]	0.3693	0.1516
GMBI [6]	0.4259	0.0997
AOD-NET [19]	0.4871	0.1667
PIX2PIX [36]	0.4148	0.1044
Ours(G Only)	0.4647	0.1161
Ours	0.4465	0.1018

framework, we also report results of the generator-only version of our method (marked as "Ours(G Only)") as an ablation experiment. Our generator-only version gave similar results to that of PIX2PIX due to the similar network structure. With our proposed loosely-coupled networks, all of the smoke is removed. The estimated smoke mask can correctly predicts the real surgical smoke most of the time, but sometimes it can fail such as the I_{middle}. The differences between our generator-only version and our final version have proved that our smoke removal network is based on the predicted smoke mask, and the combination of the smoke detection with the smoke removal can narrow the gap between simulation and reality, thus improving the overall de-smoking performance for the *in-vivo* dataset.

As there are no ground-truth smoke-free image pairs from the *in-vivo* datasets for quantitative evaluations, we adopt the Fog Aware Density Evaluator (FADE) [33] for the reference of perceptual smoke evaluation. FADE is a smoke prediction model based on natural scene statistics (NSS) and fog aware statistical features. The lower FADE score, the less perceptual fog, and vice versa. The quantitative evaluation results by FADE are reported in Table II. We can see that our method does not receive the lowest FADE score. This is because FADE is based on the statistics of non-fog scene features, which will always take the sharpness, contrast and saturation of the image into consideration. However, our learning based method is trained and focused on recovering the natural and realistic smoke-free surgical images without the emphasize on the image visual quality metrics such as sharpness, contrast and saturation. For the GAN-based method, from the previous experiments we already know that it will create some fake features (such as scars) on the images to make it look like smoke free image (that usually have high sharpness), so that PIX2PIX scores higher than our method. However, our method has the lowest std. value.

Fig. 8. The quantitative results on *in-vivo* dataset. Each column of the image matrix shows the de-smoking results from different methods, and each row shows the de-smoking results of different image samples for the same method.

V. DISCUSSION

A. Prevent Overfitting

One of the novelties of our work is that we do not require ground truth data (the smoke and smoke-free image pairs) and can achieve unsupervised training from the perspective of data requirements. The *in-vivo* experiment proves that our networks, although trained on synthetic data, can detect and remove smoke on real surgical datasets. The use of synthetic datasets for network training compensates for the lack of training datasets for medical applications, bridging a significant gap between simulation and reality. This is due to the fact that we have developed a set of techniques to prevent our networks from overfitting to the synthetic data. For example, our training data is carefully selected and rendered. The backgrounds are extracted from 91 different laparoscopic and endoscopic videos with different surgical procedures with different image colours and tones and under the presence of different surgical instruments. The smoke is rendered by a cinematic rendering engine by using random intensities, densities, textures and positions. We believe that the decomposition of the de-smoking task into two separate tasks (the smoke detection and removal) also helps to prevent overfitting. As we are not directly creating the mapping from the smoke image to the smoke-free image, but rather, we detect the area and the intensity of the smoke first, and then try to recover the smoke-free image based on the smoke prior. The use of a shallow network and drop-out for smoke detection is intentional to prevent overfitting. This solves the challenging problem that deep learning requires large amounts of hand-labeled ground truth training data, especially for medical datasets where professional knowledge is vital in the labeling

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2019.2953717, IEEE Transactions on Medical Imaging

process. Also, during the design of our training flow, we found it interesting that training the Detector and Generator together will be less over-fitted to the training set than training them separately and sequentially (train Detector first and then use Detector to train Generator). Our explanation to this is that if D and G were trained separately and sequentially, the G will be totally reliant on the precise output of D, which will lose some generalization ability to different types/qualities of smoke masks.

B. Safety Issues

During the discussion with many medical practitioners, some concerns arose about the potential risk of removing surgical smoke from images as it might confuse surgeons. In some circumstances, although smoke may block the view, it can also be a good signal for an on-going ablation process. These concerns inspired us to add the smoke detection network that provides an extra pixel-level smoke detection before the smoke removal network removes the smoke. The predicted smoke can directly be shown to surgeons or transferred to a different format for surgeons to perceive it without distraction (see potential applications in the next section).

It is also worth noting that, although a GAN framework (such as PIX2PIX) is a very good method for generating images, it can be very dangerous and care must be taken if used in medical applications due to its uncertainty and uncontrollable nature. During our experiments, we found that the GAN-based method can create fake "scars" or "reflections" to make the images look like a smoke-free image, which is totally unacceptable and may cause serious accidents if used during surgery. Our proposed method can prevent this issue by enforcing the Detector's output to be the estimated smoke rather than a binary discriminator that produces ambiguity loss and gradients during the training of the generator.

C. Application

Based on our smoke detection and removal framework, several advanced applications can be built. One of which is related to the safety issue that we mentioned earlier that surgical smoke is a good visual cue to surgeons when an ablation is taking place. As illustrated in Figure 9, our proposed method has the potential of transforming the predicted smoke removal into a secondary image or even another format (such as sound) to alert surgeons about the on-going ablation process, whilst watching the real-time de-smoked video streams.

Also, the smoke removal is not only for surgeons but also can be used as a pre-processing step for many vision-based surgical assistance systems [37] to improve the robustness to smoke.

D. Future Work

In future work, we are going to combine CNN's with the recurrent neural networks (RNN) for video sequence smoke removal. Since during surgical ablation, smoke density rises with time. RNN can help to memorize the features (such as tissue colours) when there is light smoke and have the potential to recover the features even with very high smoke densities. It will also be interesting to see whether training networks from synthetic datasets can be extended to many other tasks such as laparoscopic camera tracking, surgical instruments detection and tissue/organ segmentation, which will overcome the shortage of medical ground-truth data and greatly benefit the deep learning technology to be used in surgical scenes.

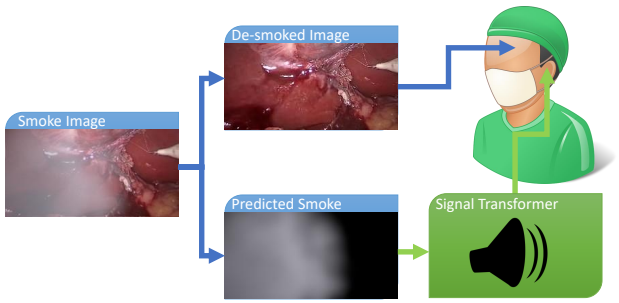

Fig. 9. Potential application of our system: transforming smoke into sound

VI. Conclusion

In this paper, we present a novel deep learning framework for real-time surgical smoke detection and removal during minimally invasive surgery. Our unsupervised training framework only needs laparoscopic images as input without a large number of hand-labeled datasets. A 3D render engine is used to randomly render smoke onto laparoscopic images to synthesize datasets for training. The novelty of this work lies in our GCN training framework that has used the smoke detection network as prior knowledge and also as the supervision for our smoke removal network. With this initiative, not only can it achieve pixel-level smoke detection, but also helps to improve the smoke removal performance compared to the state-of-the-art smoke removal methods. Our framework also yields the extra benefit of preventing over-fitting of networks to synthetic datasets, and also has many potential applications for surgical human-computer interactions.

References

[1] C. Tsui, R. Klein, and M. Garabrant, "Minimally invasive surgery: national trends in adoption and future directions for hospital strategy." Surgical Endoscopy, vol. 27, pp. 2253–2257, Jul. 2013.

[2] K. J. Weld, S. Dryer, C. D. Ames, K. Cho, C. Hogan, M. Lee, P. Biswas, and J. Landman, "Analysis of surgical smoke produced by various energy-based instruments and effect on laparoscopic visibility." Journal of endourology, vol. 21, pp. 347–351, Mar. 2007.

[3] R. Plantefève, I. Peterlik, N. Haouchine, and S. Cotin, "Patient-specific biomechanical modeling for guidance during minimally-invasive hepatic surgery." Ann Biomed Eng, vol. 44, no. 1, pp. 139–153, Jan 2016. [Online]. Available: http://dx.doi.org/10.1007/s10439-015-1419-z

[4] B. C. Ulmer, "The hazards of surgical smoke," AORN Journal, vol. 87, no. 4, pp. 721–738, apr 2008.

[5] A. Kotwal, R. Bhalodia, and S. P. Awate, "Joint desmoking and denoising of laparoscopy images," in 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), April 2016, pp. 1050–1054.

[6] A. Baid, A. Kotwal, R. Bhalodia, S. N. Merchant, and S. P. Awate, "Joint desmoking, specularity removal, and denoising of laparoscopy images via graphical models and bayesian inference," in 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), April 2017, pp. 732–736.

[7] X. Luo, A. J. McLeod, S. E. Pautler, C. M. Schlachta, and T. M. Peters, "Vision-based surgical field defogging," IEEE Transactions on Medical Imaging, vol. 36, no. 10, pp. 2021–2030, oct 2017.

[8] K. Tchaka, V. M. Pawar, and D. Stoyanov, "Chromaticity based smoke removal in endoscopic images," in Medical Imaging 2017: Image Processing, M. A. Styner and E. D. Angelini, Eds., vol. 10133, International Society for Optics and Photonics. SPIE, 2017, pp. 463 – 470. [Online]. Available: https://doi.org/10.1117/12.2254622

[9] C. Wang, F. A. Cheikh, M. Kaaniche, and O. J. Elle, "A smoke removal method for laparoscopic images," CoRR, vol. abs/1803.08410, 2018. [Online]. Available: http://arxiv.org/abs/1803.08410

[10] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An end-to-end system for single image haze removal," IEEE Transactions on Image Processing, vol. 25, no. 11, pp. 5187–5198, nov 2016.

[11] E. J. McCartney and F. F. Hall, "Optics of the atmosphere: Scattering by molecules and particles," Physics Today, vol. 30, no. 5, pp. 76–77, may 1977.

[12] S. Narasimhan and S. Nayar, "Contrast restoration of weather degraded images," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 6, pp. 713–724, jun 2003.

[13] S. K. Nayar and S. G. Narasimhan, "Vision in bad weather," in Proceedings of the Seventh IEEE International Conference on Computer Vision, vol. 2, Sep. 1999, pp. 820–827 vol.2.

[14] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 12, pp. 2341–2353, dec 2011.

[15] R. Fattal, "Single image dehazing," ACM Transactions on Graphics, vol. 27, no. 3, p. 1, aug 2008.

[16] K. Nishino, L. Kratz, and S. Lombardi, "Bayesian defogging," International Journal of Computer Vision, vol. 98, no. 3, pp. 263–278, nov 2011.

[17] R. T. Tan, "Visibility in bad weather from a single image," in 2008 IEEE Conference on Computer Vision and Pattern Recognition, June 2008, pp. 1–8.

[18] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in 2013 IEEE International Conference on Computer Vision, Dec 2013, pp. 617–624.

[19] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," in 2017 IEEE International Conference on Computer Vision (ICCV), Oct 2017, pp. 4780–4788.

[20] L. Chen, W. Tang, and W. John, "Unsupervised learning of surgical smoke removal from simulation," in Hamlyn Symposium on Medical Robotics, 2018, pp. 75–76.

[21] C. Wang, A. K. Mohammed, F. A. Cheikh, A. Beghdadi, and O. J. Elle, "Multiscale deep desmoking for laparoscopic surgery," in Medical Imaging 2019: Image Processing, vol. 10949, 2019. [Online]. Available: https://doi.org/10.1117/12.2507822

[22] K. Tang, J. Yang, and J. Wang, "Investigating haze-relevant features in a learning framework for image dehazing," in 2014 IEEE Conference on Computer Vision and Pattern Recognition, June 2014, pp. 2995–3002.

[23] S. Bolkar, C. Wang, F. A. Cheikh, and S. Yildirim, "Deep smoke removal from minimally invasive surgery videos," in 2018 25th IEEE International Conference on Image Processing (ICIP), Oct 2018, pp. 3403–3407.

[24] M. Ye, E. Johns, A. Handa, L. Zhang, P. Pratt, and G.-Z. Yang, "Self-supervised siamese learning on stereo image pairs for depth estimation in robotic surgery," in Hamlyn Symposium on Medical Robotics, 2017, pp. 27–28.

[25] A. P. Twinanda, S. Shehata, D. Mutter, J. Marescaux, M. de Mathelin, and N. Padoy, "EndoNet: A deep architecture for recognition tasks on laparoscopic videos," IEEE Transactions on Medical Imaging, vol. 36, no. 1, pp. 86–97, jan 2017.

[26] R. Bansal, G. Raj, and T. Choudhury, "Blur image detection using laplacian operator and open-cv," in 2016 International Conference System Modeling Advancement in Research Trends (SMART), Nov 2016, pp. 63–67.

[27] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., 2015, pp. 234–241.

[28] A. P. Twinanda, S. Shehata, D. Mutter, J. Marescaux, M. de Mathelin, and N. Padoy, "Endonet: A deep architecture for recognition tasks on laparoscopic videos," IEEE Transactions on Medical Imaging, vol. 36, no. 1, pp. 86–97, Jan 2017.

[29] R. Stauder, D. Ostler, M. Kranzfelder, S. Koller, H. Feußner, and N. Navab, "The TUM lapchole dataset for the M2CAI 2016 workflow challenge," CoRR, vol. abs/1610.09278, 2016. [Online]. Available: http://arxiv.org/abs/1610.09278

[30] A. Galdran, J. Vazquez-Corral, D. Pardo, and M. Bertalmio, "Fusion-based variational image dehazing," IEEE Signal Processing Letters, pp. 1–1, 2016.

[31] M. Sulami, I. Glatzer, R. Fattal, and M. Werman, "Automatic recovery of the atmospheric light in hazy images," in 2014 IEEE International Conference on Computational Photography (ICCP), May 2014, pp. 1–11.

[32] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," IEEE Transactions on Image Processing, vol. 24, no. 11, pp. 3522–3533, nov 2015.

[33] L. K. Choi, J. You, and A. C. Bovik, "Referenceless prediction of perceptual fog density and perceptual image defogging," IEEE Transactions on Image Processing, vol. 24, no. 11, pp. 3888–3901, nov 2015.

[34] A. Galdran, J. Vazquez-Corral, D. Pardo, and M. Bertalmío, "Enhanced variational image dehazing," SIAM Journal on Imaging Sciences, vol. 8, no. 3, pp. 1519–1546, jan 2015.

[35] D. Berman, T. Treibitz, and S. Avidan, "Air-light estimation using haze-lines," in 2017 IEEE International Conference on Computational Photography (ICCP), May 2017, pp. 1–9.

[36] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017, pp. 5967–5976.

[37] L. Chen, W. Tang, N. W. John, T. R. Wan, and J. J. Zhang, "Slam-based dense surface reconstruction in monocular minimally invasive surgery and its application to augmented reality," Computer Methods and Programs in Biomedicine, vol. 158, pp. 135 – 146, 2018.